this view, only the *leu*-3 cistron need be sensitive to the regulatory control of the inducer α-IPM, but some signal generated as a consequence of *leu*-3 activity would be required for induction of the isomerase and dehydrogenase synthesis and proper derepression of synthetase synthesis.

While the notion of the involvement of a specific locus in the coordination of function of several physiologically related but unlinked cistrons is at least suggested by the information presently available, it by no means simplifies conceptually the complexity of the regulatory mechanism involved. The physiological and genetic organization of the enteric bacteria and the fungi differs radically, but no specific aspect of the enzymology or genetics of the two groups of organisms suggests why in any specific instance a genetically disperse multisignal system of control should be more advantageous than the regulation of function by a set of linked cistrons responding to the intensity of a single signal. Clearly the extensive linkage displayed by the *arom* cistrons[13] indicates that *Neurospora* probably makes use of both single and multisignal mechanisms of regulation. It is exactly this diversity that seems central to the general problem of the relation between regulation and function.

These investigations could not have been accomplished without the skillful technical assistance of Mrs. Evelyn Gilmore and Mrs. Geraldine Williams.

* This research was supported by National Science Foundation grant GB-727 and U.S. Public Health Service grant GM-07250.

[1] Margolin, P., *Genetics*, **48**, 441 (1963).

[2] Lacroute, F., *Compt. Rend.*, **258**, 2884 (1964).

[3] Webster, R. E., and S. R. Gross, *Biochemistry*, in press; and Webster, R. E., C. A. Nelson, and S. R. Gross, *Biochemistry*, in press.

[4] Gross, S. R., R. O. Burns, and H. E. Umbarger, *Biochemistry*, **2**, 1046 (1963).

[5] Burns, R. O., H. E. Umbarger, and S. R. Gross, *Biochemistry*, **2**, 1053 (1963).

[6] Lester, H. E., and S. R. Gross, *Science*, **129**, 572 (1959).

[7] Friedemann, T. E., and G. E. Haugen, *J. Biol. Chem.*, **147**, 415 (1943).

[8] Rennert, O. M., and H. S. Anker, *Biochemistry*, **2**, 471 (1963).

[9] Jacob, F., and J. Monod, in *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 26 (1961), p. 193.

[10] Gross, S. R., C. Jungwirth, and E. Umbarger, *Biochem. Biophys. Res. Commun.*, **7**, 5 (1962).

[11] Schlesinger, S., and B. Magasanik, *J. Mol. Biol.*, **9**, 670 (1964).

[12] Gross, S. R., these PROCEEDINGS, **48**, 922 (1962).

[13] Gross, S. R., and A. Fein, *Genetics*, **45**, 885 (1960); and Giles, N. H., M. E. Case, and C. W. H. Partridge, *Genetics*, **52**, 444 (1965).

---

*ON THE EVOLUTION OF THE GENETIC CODE**

BY C. R. WOESE

DEPARTMENT OF MICROBIOLOGY, UNIVERSITY OF ILLINOIS, URBANA

Communicated by S. Spiegelman, October 8, 1965

Now that the cryptographic aspect of the genetic code has been solved in essence, it has become apparent that we must undertake to answer the question of how this highly ordered array of codon assignments came into being. One aspect of the order possessed by the codon catalogue is approximated by a general type of code,

called an "$a \times b \times c$ code," predicted on theoretical grounds some time ago.[1, 2] At the time that this theoretical code was proposed, it was considered to be derivable from one of several general types of mechanisms; one possibility was a steric interaction between amino acids and oligonucleotides. Another postulated an interaction between codon and sRNA such that one kind of sRNA recognized *more than* one kind of codon.[2, 3] However, it does *not* appear necessary to invoke *any* mechanism, be it amino acid-oligonucleotide steric interaction or particular kinds of codon-sRNA interactions, etc., to account for the high degree of order shown by the codon catalogue; Sonneborn has argued rather convincingly that a purely stochastic process, involving selection during evolution to *minimize* the lethal effect of ordinary mutations, could account for the kind of order the codon catalogue manifests.[4]

In addition to possessing the strong intercodon order of an "$a \times b \times c$" code, the codon catalogue has more recently been shown to manifest very definite correlations among the codon assignments for *"related"* amino acids.[5] This latter constraint is *not* predicted on the basis of the simple sRNA-codon degeneracy mechanism above, but the Sonneborn stochastic model *is* compatible with such a constraint. However, rather than analyzing in detail these various models ostensibly accounting for the order in the codon catalogue, I should like at this time to present an alternative explanation for how a codon catalogue of the type observed could have evolved—an hypothesis suggested by the resemblance of the kinds of errors characterizing the translation process to the type of order possessed by the codon catalogue.[6] In brief, the codon catalogue which we observe today is considered to have arisen through a series of evolutionary steps which served gradually to reduce an initial inherent high error rate in the translation process of the primitive cell.

The Characteristics of Translation Errors.—It is now well known that high rates of error can be produced artificially in the translation process *in vitro* by any of a number of treatments which appear to amount to creating suboptimal conditions—e.g., high Mg ion level, high pH, low temperature, presence of streptomycin, etc.[7–9] In addition, "errors" also result when synthetic mRNA's containing unnatural analogues of the normal bases are translated by *in vitro* systems.[10] Further translation errors apparently similar to those produced *in vitro* can be observed *in vivo*—e.g., in streptomycin-treated cells or in cells possessing particular mutations affecting the ribosome itself.[11]

The first important fact to recognize about translation errors is that for a given mRNA (poly U has been the main one studied) the errors characterizing it are to a first approximation the same, regardless of which of the "suboptimal conditions" are used to bring them about, including the use of mRNA's containing abnormal analogues of the natural bases. This is readily seen in the examples of translation errors presented in Table 1. The common pattern that appears to characterize the translation errors for the case of the codon UUU is:[12]

Position on the codon	I	II	III
Base mistaken for U	C,A	C-low (A-very low)	C,A,G

Also under "normal" conditions for the *in vitro* system—i.e., 0.015 *M* Mg ion, 37C°—

TABLE 1

Amino acid	Poly U Sm[9] (or DHSm)	Poly U DHSm subopt. temp.[8]	Poly BrU[10]	Poly BrC[10]
Phe	100	100	100	—
Leu	20	80	180	—
Ilu	60	100	490	—
Tyr	0	1.5	10	—
Ser	7	9	40	0
Val	—	0.5	0	—
Cys	—	—	0	—
Pro	—	—	—	100
Thr	—	—	—	80
His	—	—	—	0
Gln	—	—	—	4

Translation errors caused by various suboptimal conditions in the *in vitro* system. All incorporations are relative to phe incorporation which is taken as 100. The values of incorporations due to translation errors are corrected for any error incorporation occurring in the absence of the agent used to produce error. Sm, streptomycin; DHSm, dehydrostreptomycin.

the error rate in the III position of the codon is about 100 times that for the I position, which in turn appears to be 10-fold greater than that for the II position.[7] This type of error pattern seems not to be confined to the codon UUU, but perhaps embraces all those codons containing pyrimidines in the II position: the parallels between results obtained using poly BrU and poly BrC, as well as the types of errors characterizing a UG copolymer, support some generalization of this sort.[8] In contrast, however, it is apparent that errors involving the *purine*-rich codons (perhaps the II purine codons) follow a very different pattern, as the more recent data of Davies *et al.* show.[13] In particular, the over-all error rate is considerably lower for the high purine codons than for the high pyridine ones, and there is no readily discernible pattern to the former as there is with the latter.

The second important fact about these translation errors, at least those involving the pyrimidine-rich codons, is that their pattern bears a striking similarity to the order manifested by the codon catalogue:[14] the III position in the codon, the most error-prone, is also the one manifesting practically all the degeneracy in the codon catalogue; the second most error-prone codon position, namely I, is that position which defines the codon assignments for "related" amino acids;[5] position II in the codon, the least error-prone, is also the one manifesting neither a "related" amino acid constraint, nor any base "equivalence."

The Translation Error (TE) Model for Evolution of the Codon Catalogue.—The translation mechanism employed by the cells of today is a complex hierarchy of macromolecules which functions with speed and great accuracy. It is self-evident that such a hierarchy is the product of a complex evolutionary process, which in turn makes it essentially certain that at some stage sufficiently early in evolution the translation mechanism was a far more rudimentary thing than at present, in particular far more prone to make translation errors.[15] (The reasoning behind this will become clearer as the argument proceeds.) Just what sort of mechanism was involved in translation at these early times we shall leave open for the present—it is not necessary to invoke sRNA or the ribosome as we now know them. However, what we shall assume about a very primitive translation system is that the mechanism was such that one could consider particular amino acids to be "assigned" to particular codons—it seems likely that *ambiguous* codon assignments would have been common, however. We further assume the mechanism was such that errors in translation were *extreme*—to such an extent that the probability of translating

any given gene (or mRNA) entirely correctly was essentially *zero*. Finally, it is assumed that the pattern of errors characterizing the early translation process was qualitatively like the above-described one, known to exist today.

Two important consequences stem immediately from this concept of error-ridden translation in the primitive cell: (1) since perfect translations of a gene are negligible (and thus *no* two proteins in the cell are identical), the proteins produced by any given gene will have to be what we shall call "statistical proteins"—i.e., to each gene there corresponds a *group* of proteins whose primary structures are related to some theoretical average primary structure, which in turn characterizes the gene; (2) it would be relatively easy to alter actual codon assignments, since this would in many cases have little or no deleterious effect on the already rather chaotic situation existing in the primitive cell.

"How could such a cell contain any enzymes at all, and so be visible?" We cannot answer this definitely at the present state of our knowledge. Nevertheless, it is essentially a certainty that at an early enough stage in evolution such cells as these did exist and some had to be viable. It might help in rationalizing this if we point out: (1) the environment of such cells was possibly very rich (the "Oparin ocean"), and thus few or no enzymes of intermediary metabolism would be required; (2) our concept of what constitutes an enzyme is probably very biased by the fast-acting, high-specificity enzymes we observe in cells today—these being the end product of an extensive evolution; and (3) even random-sequence polypeptides have been shown to manifest low levels of certain catalytic activities.[16] Therefore, we shall assume that very primitive cells possessed a few kinds of rudimentary, low-specificity, slow-acting "enzymes" derived from various of their "statistical proteins," though granted it is also likely that such cells contained a *large* fraction of proteins which were completely useless enzymatically (or perhaps in some cases even harmful).

It is quite evident that the main problem in evolution for this early cell was to reduce errors in the translation process—eventually switching from reliance on "statistical proteins" to proteins of reasonably well-defined primary structures. It is highly *unlikely* that this could have been accomplished merely by "evolving a more efficient translation apparatus." The reasoning here can best be understood if we argue the point in reverse. If a (modern) cell makes mistakes in translation—above a certain level—*and* some of the faulty proteins so produced are destined themselves to become a part of a translation apparatus, say a ribosome, then the new ribosome could be *more* prone to make translation mistakes than its predecessors, which in turn would lead to synthesis of similarly faulty ribosomes at an even faster rate, until finally a catastrophic situation prevails, in which the cell line translates in a completely error-ridden fashion (and so becomes nonviable).[17] What we may be witnessing in evolution of the translation process is something like the reverse of this. The primitive cell was faced with the seeming paradox that in order to develop a more accurate translation apparatus, it had first to translate more accurately. The way out of this paradox, of course, is that although unable *really* to reduce the translation error rate, the primitive cell can do something tantamount to this by adjusting the codon catalogue so that the *effect* of the translation errors is lessened. Let us try to picture this early evolution in a bit more detail.

The very primitive cell, relying on "statistical proteins" for its rudimentary en-

zymes, probably did not "recognize" 20 individual amino acids, although 20 or thereabouts could have been "assigned" to codons. What this cell did recognize were *groups* of amino acids. There could be no conceivable advantage for the cell to distinguish leu from val or ilu, asp from glu, lys from arg, etc., at this stage. In fact, the amino acids used today can be divided crudely into two groups, the relatively "nonfunctional" amino acids, like phe, leu, val, ala, thr, etc., and the relatively "functional" ones, such as tyr, his, lys, glu, try, etc. The early cell would then perhaps work in terms of a very few amino acid *groups*—the "nonfunctional" group plus a few different kinds of "functional" groups. To the degree that the early cell used its "statistical proteins" as enzymes, it relied most heavily upon the placement, number, and kind of "functional" amino acids—the "nonfunctional" ones serving mainly as "spacers" or to provide an hydrophobic environment. Therefore, if the primitive cell started with a completely random, highly ambiguous set of codon assignments, the first step in improving the translation efficiency would be the gravitation of the codon assignments for the "functional" groups of amino acids toward the *least* error-prone codons, say the II purine codons, leaving the *more* error-prone, II pyrimidine codons, for the large group of "nonfunctional" amino acids. Once translation was improved to some extent by this maneuver, it might eventually become feasible for the cell to make finer distinctions within amino acid groups, dividing them into subgroups and so on, until finally, individual amino acids would be recognized as such. As the end result of this sort of evolution there would exist a codon catalogue of the form we see today, one which can reduce effective translation errors to a minimum. This is a catalogue with the following two properties: the probability is maximal that (1) a mistake in reading the underlying codon will still lead to no change in the overlying amino acid—both the actual codon and the one mistaken for it being assigned to the *same* amino acid, and (2) if a mistake in reading the underlying codon does result in introducing the incorrect amino acid into a polypeptide, then the incorrect amino acid will be, on the average, as closely "related" to the intended one as possible.

It can next be argued that only when the codon catalogue had evolved to a stage of (nearly) optimal error reduction would it then be *possible* for the cell to translate with sufficient accuracy to begin evolving a superior type of translation *mechanism*.[18] The advent of this superior type of translation apparatus can be viewed as a major turning point in evolution, for it undoubtedly made possible the emergence of cells as we know them today. The ultimate (modern) translation apparatus conceivably is not only vastly improved from the point of view of translation error frequency, but also functions much more rapidly than its predecessors. This step in evolution was indeed a "supermutation" which gave to cells possessing the new translation apparatus such a selective advantage that cells not having it were soon eliminated from the face of the earth.

Discussion and Summary.—The TE model for evolution of the genetic code starts with a primitive cell possessing *random, ambiguous* codon assignments, and a *very* error-ridden translation process, and shows how such cells, from the "necessity" of minimizing the effects of translation errors, can evolve the highly ordered codon catalogue we observe today. The real value of this model, however, is not in evolving the code according to a particular scheme. It is in the recognition that at sufficiently early stages in evolution the fundamental information-transferring

processes, i.e., translation, replication, and transcription, must have been error-ridden, and in the derivation of some of the possible consequences of this fact. (This concept and its general consequences are not restricted to codes derived in basically stochastic ways.) Let us discuss some of these consequences in a bit more detail.

All of evolution, with the possible exception of the more recent evolution—i.e., the last billion years or thereabouts—must be viewed as being limited and therefore defined by the accuracy with which information transfer can take place in the cell. Very probably the evolution of accurate translation mechanisms, etc., occurred in a series of more or less discrete stages, each beginning with an improvement in a given information transfer process (through mutation), which then lead to the gradual evolutionary working out of its ramifications, which, in turn, set the stage for—made possible—a further improvement in information transfer, etc. Evolution during any of these hypothetical stages should be *qualitatively* different from that occurring in any other stage, for the basic cell type would probably differ from one stage to another more drastically than do any of the cell types now extant. If we picture an early cell as possessing solely "statistical proteins," from which it can fashion only a very few low-specificity, slow-acting, inaccurate "ur-enzymes," it is quite clear that such a cell is not capable of evolving the great variety of cell types found on earth today—an evolution which requires a cell to possess many precisely defined enzyme functions, capable of subtle modifications, and a metabolic pattern tightly interrelated through various feedback controls of its enzymes.[19] Thus to date, we may have witnessed only the final "divergent" stage in evolution. All the previous stages—which might be considered "convergent" by virtue of their having as the sole or main "goal" the improvement of some feature of information transfer—have perhaps gone undetected so far.

The author is very grateful to Dr. S. Spiegelman for discussions of this topic and criticism of the manuscript.

* This work was supported by NSF grant GB-2228.

[1] Woese, C., *Nature*, **194**, 114 (1962).

[2] Woese, C., *ICSU Rev.*, **5**, 210 (1963).

[3] This latter mechanism did not seem too likely a few years ago in view of the demonstration that leu had at least three *different* sRNA's for three of its codons. However, more recent evidence suggesting that perhaps not *all* codons have separate sRNA's has led to the re-emergence of this mechanism in a more detailed form—Crick, F. H. C., *The Wobble Hypothesis*, unpublished.

[4] Sonneborn, T. M., in *Evolving Genes and Proteins*, ed. V. Bryson and H. J. Vogel, (New York: Academic Press, 1965).

[5] Woese, C., these PROCEEDINGS, **54**, 71 (1965).

[6] Woese, C., "The Genetic Code-1964," in *Advances in Theoretical Biophysics*, ed. A. Cole (New York: Marcel Dekker, in press).

[7] Szer, W., and S. Ochoa, *J. Mol. Biol.*, **8**, 823 (1964); Grunberg-Manago, M., unpublished results.

[8] Friedman, M., and I. B. Weinstein, these PROCEEDINGS, **52**, 988 (1964).

[9] Davies, J., W. Gilbert, and L. Gorini, these PROCEEDINGS, **51**, 883 (1964).

[10] Grunberg-Manago, M., and A. M. Michelson, *Biochim. Biophys. Acta*, **80**, 431 (1964).

[11] Gorini, L., and E. Kataja, these PROCEEDINGS, **51**, 995 (1964).

[12] This pattern has been deduced from the following facts: val-GUU and cys-UGU are *not* generally incorporated by error when a poly U or "poly U-like" message is used, while ilu-AUU, and leu-CUU, UUA, UUG are readily incorporated as mistakes under the same conditions. Ser-

UCU is incorporated at a lower level, and tyr-UAU is sometimes incorporated at a still lower level. Further, von Ehrenstein, by using isolated sRNA's in an *in vitro* protein-synthesizing system has shown val, tyr, and cys sRNA's not to respond to poly U; whereas ilu, ser, and the *three* leu sRNA's which respond normally to poly UC, poly UA, and poly UG, respectively, *all* do respond to poly U under this suboptimal condition (von Ehrenstein, unpublished results). I have assumed also that CIII can be mistakenly read as UIII on the basis of the results of Nirenberg and coworkers[20] and Söll *et al.*,[14] showing that phe sRNA appears not to distinguish between the two phe codons UUU and UUC. Of course, this may not be an error at all, but as discussed above, may represent an actual recognition of two codons by one sRNA—which for the present purposes we can consider an "error," however.

[13] Davies, J., L. Gorini, and B. Davis, *J. Mol. Pharmacol.*, **1**, 93 (1965).

[14] Söll, D., E. Ohtsuka, D. S. Jones, R. Lohrmann, H. Hayatsu, S. Nishimura, and H. G. Khorana, these PROCEEDINGS, **54**, 1378 (1965); Nirenberg, M. W., P. Leder, M. Bernfield, R. Brimacombe, J. Trupin, F. Rottman, and C. O'Neal, these PROCEEDINGS, **53**, 1161 (1965).

[15] We shall confine our attention at present to the effect of *translation* errors, realizing, of course, that high rates of error could and probably did exist at the same time for the transcription and/or replication processes. The interrelationship between the latter type of error and translation errors will be discussed elsewhere.

[16] Fox, S., *Bioscience*, **14**, 13 (1964).

[17] Orgel, L. E., these PROCEEDINGS, **49**, 517 (1963).

[18] It must be appreciated what relatively profound effects (selective advantage) small increases in the translation efficiency could have. The relation of cause to effect here—cause being the decreased translation error frequency, effect being the "improved" cell function— is an "*n*th power" sort of phenomenon. The probability of making a perfect translation of a given mRNA is approximately $(1 - E)^n$, where E is the average probability of making an error in reading a codon, and n is the number of codons in the mRNA. Therefore, if E is initially rather large, small changes in it will have a drastic effect on the function $(1 - E)^n$. Thus, adjustments in codon assignments which produce only slight improvements in translation error frequency in these error-ridden cells could still have very profound effects in terms of improved cell function, in terms of selective advantage. The feedback between improved translation and fidelity of transcription and/or replication would also play a role here.

[19] The name "ur-enzymes" has been suggested by K. C. Atwood to describe enzymes such as duplicases, transcriptase, etc.—in general, the tape-reading systems—absolutely essential to any cell.

[20] Leder, P., and M. W. Nirenberg, these PROCEEDINGS, **52**, 1521 (1964); Bernfield, M. R., and M. W. Nirenberg, *Science*, **147**, 479 (1965).

---

*EFFECT OF THE INTERVAL BETWEEN IRRADIATION AND CONCEPTION ON MUTATION FREQUENCY IN FEMALE MICE**

BY W. L. RUSSELL

BIOLOGY DIVISION, OAK RIDGE NATIONAL LABORATORY, OAK RIDGE, TENNESSEE

Communicated by Alexander Hollaender, October 19, 1965

Determinations of the effect of germ-cell stage on induced mutation frequency have been very limited in female mammals, owing to the fact that permanent sterility sets in quickly after exposure to the doses and dose rates of X and gamma radiation that are necessary for an adequate yield of mutations. Thus, accurate measurements of oöcyte mutation frequencies have been made only in mature and nearly mature follicle stages. The adult mouse contains no germ cells of developmental stages preceding the dictyate oöcyte, and, although this nuclear state persists